ISRIC Digital Data Preservation Policy

Summary: This Digital Preservation Policy describes the policy of ISRIC - World Soil Information (hereinafter referred to as ISRIC) with respect to the preservation of its digital soil resources (*sensu* datasets). It consists of the following sections: 1) Introduction, 2) Purpose, 3) Organisational strategy alignment, 4) Scope, 5) Policy principles, 6) Standards, 7) Sustainability and preservation planning, 8) Continuous improvement, 9) Relationship to other policies and documents, and 10) Policy review date. Section 8 includes a self-assessment, following the community-based Digital Preservation Coalition Rapid Assessment Model (DPC RAM), aimed at setting goals for further development and to monitor progress. This policy does not exist in isolation; other corporate policies which cover our daily functioning as World Data Centre for Soils are listed in Section 9.

Last updated: 1 July 2021 [doi: 10.17027/isric-wdcsoils.202102]

1. Introduction

This policy outlines the framework for a digital preservation policy at ISRIC – World Soil Information and summarises its purpose, scope, and objectives. ISRIC, legally registered as International Soil Reference and Information Centre, will ensure the long-term continuity of digital assets in line with its role as regular member of the ISC-WDS (International Science Council, World Data System).

Mission: As the custodian of global soil information we produce, gather, compile and serve quality-assessed soil information together with our partners at global, national and regional levels. We stimulate the use of this information to address global challenges through capacity building, awareness raising and direct cooperation with users and clients. ISRIC is a trusted digital repository holding a CoreTrustSeal (CTS) certification and accredited as a regular member of the ISC World Data System; in that capacity it is known as WDC-Soils. ISRIC provides data to the international community, in accord with the license specified by each data provider[1].

This policy helps ISRIC meet its legislative and accountability requirements and the expectations of its user community concerning the preservation of soil-related datasets[2]. Excluded are scanned maps and reports held in the ISRIC Soil Library and Map Collection[3] that are preserved under a service level agreement (SLA)[4] with the Wageningen UR library.

---

[1] https://www.isric.org/about/data-policy
[2] This policy has been adapted from the BGS, DANS and 4TU.ResearchData policies
[3] https://www.isric.org/explore/library
[4] https://www.isric.org/overview-service-level-agreements

Corporate administrative records, which are covered by procedures described in ISRIC's Operational Handbook (for internal use only), are not considered in this policy.

2.   Purpose

The aim of the ISRIC digital preservation programme is to ensure the longevity of the digital information assets of the organisation in a sustainable way by addressing the factors which risk making them unusable and inaccessible. It demonstrates that ISRIC as part of the wider digital preservation community is aware of the need to comply with relevant standards to ensure the user requirements of its designated user community and other stakeholders are fulfilled.

The overall approach is to develop a scalable and modular preservation programme to guide the management, usability and accessibility of digital objects (in particular soil-related datasets) selected for long-term preservation because of their scientific, informational or evidential value through the technological and other changes that will occur.

The benefits of long-term access to the soil data[5] held at ISRIC for re-use are both scientific and financial and apply equally to all users. The economic, reputational and cultural risks of failing to address long-term accessibility through digital preservation are minimised through the preservation programme.

This document includes a self-assessment following the community-based, Digital Preservation Coalition Rapid Assessment Model (DCP RAM), which details the tasks and actions to be taken to achieve the objectives of this policy.

3.   Organisational strategy alignment

ISRIC is the World Data Centre for Soils (WDC-Soils) within the ISC World Data System. Since 1966, it receives funding from the Netherlands Government[6] for its core activities[7].

Digital preservation must support and align with the broader strategic goals of ISRIC and the World Data System. These goals include:

- Becoming the premier provider of global soil information to support international science communities, policy communities and the private sector dealing with issues including food production, land and water management, climate change, environmental quality, social justice, land use planning, and biodiversity.
- Maintaining FAIR data assets: findable, accessible, interoperable and reusable data that is trusted in a way that suits our stakeholders' needs.
-

---

[5] https://data.isric.org/
[6] https://www.isric.org/about/history
[7] https://www.isric.org/about/vision-mission

- Emerging technology: the appearance of new and rapidly evolving technologies generates opportunities for new scientific discovery.
- Scientific e-infrastructure: develop spatial data infrastructure (SDI) for integration with the broader international scientific community .
- Digital environment: Combine data across disciplines to generate new insight, using high-performance computing to create soil property maps and a range of derived products (e.g. assessments of land degradation, land suitability, soil fertility and hydraulic characteristics and soil functions).
- Best environment for research and innovation: Maintain and enhance ISRIC expertise, infrastructure, services and facilities, which benefit the global environmental science community and global partners. Ensure a pipeline of talent with the skills to resolve future challenges and take a whole-systems approach. Engage with the public and stakeholders to draw-in expertise and share knowledge through our community of practice programme.

4. Scope

The scope of this policy is limited to the organisation's digital data holdings and applies to born-digital and digitised soil data. It covers all research data and associated metadata in all formats accessible from ISRIC's Soil Data Hub[8]:

- ISRIC owned data
- Data generated at ISRIC using shared source data (in accord with the licence specified by the data provider)
- Third party and/or co-generated data (in accord with licence(s) specified by each data provider)

5. Policy principles

This policy follows main principles of digital curation:

Authenticity: The data is what it purports to be, is created or sent by the purported person, and at the purported time. This is shown e.g. in the provenance of data and preservation metadata.

Integrity: The data stored in the ISRIC Repository is complete and unaltered. This is achieved by appropriate ingestion processes and using digital signatures, fixity and check sum checks (md5, under development), and persistent identifiers for the data.

Reliability: The data accurately reflects the original context of data creation and is trustworthy. This is achieved by documenting and capturing the contextual metadata, including access to the codes and primary data used to generate the derived data, and ensuring its completeness.

Usability: The data can be located, retrieved, presented and interpreted. For this, ISRIC uses preferred and open file formats, monitoring of obsolescence (software, hardware), persistent data identifiers, and digital rights management as part of metadata capture.

---

[8] https://data.isric.org/

ISRIC implements FAIR[9] data (Findability, Accessibility, Interoperability, and Reusability) principles. Through our CoreTrustSeal certification we endorse the TRUST[10] principles (Transparency, Responsibility, User Focus, Sustainability and Technology).

The following aspects are considered:

Preservation objectives: All soil-related data must be managed and maintained, backed up securely, stored and delivered to all users by ISRIC. The objectives for ISRIC with regard to provision of relevant, comprehensive and up-to-date information include:

- Creation and capture of good metadata and other documentation required to understand the data and to use it correctly.
- Data integrity, evaluation and validation according to standardised ingest procedures, providing users with the confidence that what they retrieve from the archive is what it says it is (such procedures are typically documented in peer-reviewed Journal papers as well as dataset specific FAQ-webpages).
- Professional cataloguing of data using appropriate metadata standards.
- Maintaining the authenticity, integrity and reliability of datasets stored and preserved through the inevitable changes to hardware platforms, operating systems, and software environments.
- Maintaining the long-term availability, discoverability and accessibility of datasets[11], enabling users to work with the data in the way they need to (e.g. facilities will be implemented to enable legacy data formats to be read by future generations of software).

Integration with data management: Digital preservation is an integral part of RDM (Research Data Management) activities from early on in the data lifecycle. Preservation planning in the pre-ingest and ingest stages increases efficiency, for example through implementation of Data Management Plans (DPM), will reduce costs in migrating or rescuing data and minimises reputational risk due to data becoming inaccessible. Early consideration of GDPR requirements concerning the handling of privacy-sensitive data is considered an important element in this process.

Acquisition of digital materials: ISRIC collections policy, digital strategy and science strategy guide the prioritisation of materials selected for long-term preservation. Procedures to appraise the long-term scientific value of data offered to ISRIC for consideration in its WoSIS and SoilGrids workflows, including guidance to data providers for submitting datasets, is made available on the ISRIC website.

Preservation metadata: 'Preservation metadata is metadata that supports the distinct requirements of digital preservation: maintaining the availability, identity, persistence, renderability, understandability and authenticity of digital objects over long periods of time.'[12] The preservation strategy describes the processes for preservation metadata capture and the use of permanent identifiers (UUID) and DOIs.

---

[9] http://dx.doi.org/10.1038/sdata.2016.18
[10] https://www.rd-alliance.org/trust-principles-rda-community-effort
[11] https://data.isric.org/
[12] http://dx.doi.org/10.7207/twr13-03 , p.2

Technical implementation: Details of technical implementation, such as back up procedures, file migration and   data rescue strategies are covered through SLA's[13] with Wageningen UR and documented in the ISRIC Procedures Handbook (for internal use only). Technical (and scientific) workflows for data ingestion and processing are regularly updated to accommodate rapid changes in tools, technologies and digital platforms employed.

Preservation risk assessment: The preservation strategy includes a risk assessment to identify and prioritise the most at-risk data and the actions selected to strengthen their long-term accessibility in line with available resources.

Funding and resources: The support of ISRIC senior management is required to maintain staffing, develop skills levels and secure long-term funding in order to carry out preservation activities.

Legal basis: As per its data policy, ISRIC aims to make all data openly available insofar as possible. There are safeguards in place for sensitive and commercially restricted data, including agreed embargo periods and access restrictions.

ISRIC must also have legal rights to preserve any digital content kept in its archives. The rights must include the ability to delegate a successor, should the original organisation cease operations in accordance with WDS-requirements. Any legal agreements must always remain with the archived digital content and be preserved with it. They must also be updated if there are changes in the licensing agreements.

6. Standards

The organisation will follow the broad guidance given in standards and best practice guidance to support the level of preservation required. These include:

- The *Open Archive Information System (OAIS) model (ISO 14721)* will be used as an underpinning policy model. It covers functional areas necessary for digital archives including data ingestion, archival storage, data management, administration, preservation planning, and access to data.
- *Information Security Standard (ISO 27001:2013)* balances access and security of digital information.The standard requires that an organisation understands what information assets it holds and ascertains the value of these assets.
- *ISO 19115-1:2014 Geographic information — Metadata* provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.
- *ISO 23081: Information and documentation – Managing metadata for records* includes references to preservation metadata to underpin the continued authenticity, reliability, usability and integrity of digital information to support its preservation.

---

[13] https://www.isric.org/overview-service-level-agreements

- *CoreTrustSeal* is an international, community based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures. The CoreTrustSeal Trustworthy Data Repositories Requirements reflect the characteristics of trustworthy repositories. ISRIC, WDC-Soils since 1989, has been certified since July 2018 and will renew the certification every three years to maintain and enhance its certification level.

ISRIC maintains some good practice guidance on its website, including recommended file formats as well as international standards[14] in the soil domain.

7.   Sustainability and preservation planning

Funding of the ISRIC digital preservation function is provided through core funding by the Netherlands Government.

ISRIC will put in place processes to review preservation and infrastructure requirements, to evaluate preservation strategies used, and to monitor technological changes as part of the preservation planning function. These processes will be integrated in the data lifecycle to minimise the risks and reviewed on a regular basis to ensure they remain up to date.

We will also strive to comply with the EU code of conduct for scientific integrity[15] of collaboration, excellence, innovation and integrity whilst developing and implementing our digital preservation programme. This means we will consider the environmental impact of our long-term data storage and preservation activities and embed sustainability into existing and new processes and use of resources.

8.   Continuous Improvement

ISRIC will use benchmarking schemes such as the DPC Rapid Assessment Model (RAM)[16], that are based on self-assessment, at regular intervals to measure progress (Figure 1).

---

[14] https://www.isric.org/international-soil-standards
[15] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/european-code-of-conduct-for-research-integrity_horizon_en.pdf
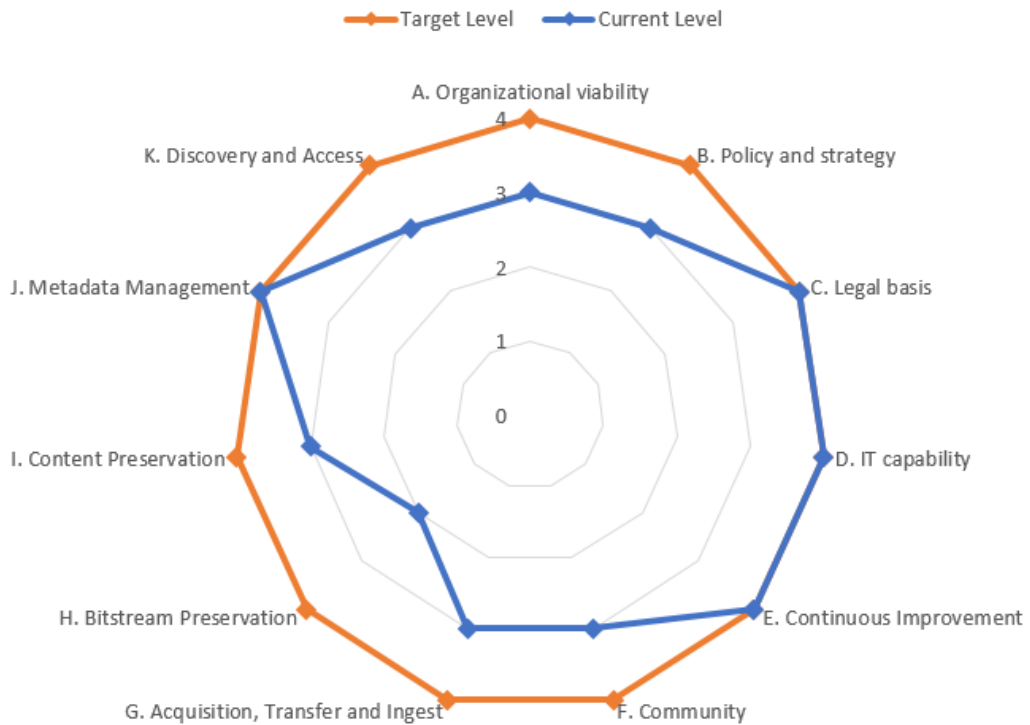[16] https://www.dpconline.org/digipres/dpc-ram

Figure 1. Digital Preservation Coalition Rapid Assessment Model (DPC RAM) self-assessment
(Current versus target level).

9.   Relationship to other policies and documents

This policy does not exist in isolation. Other corporate policies and contextual links, which cover our daily functioning as World Data Centre for Soils and have an impact on digital preservation, include:

- ISRIC Data and Software policy
- ISRIC Collection Management Policy
- ISRIC Service Level Agreements
- ISRIC Privacy Statement
- ISRIC Procedures and Standards for data deposit and ingest, including the metadata service [17] and the acceptable file formats [18] (from DANS)

In addition, the stakeholder requirements (SLAs, MOUs, formal contracts etc.) and the role of ISRIC within national and international soil-related communities will also be considered in the preservation framework.

---

[17] https://data.isric.org/

[18] https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats

10.   Document control

This policy will be reviewed on a regular basis at the instigation of the ISRIC Data Steward.

Version:	1.0
Created/Last updated:	1 July 2021
Written by:	Niels H. Batjes, Coordinator WDC-Soils
Approved by:	Rik van den Bosch, Director ISRIC
Review date:	1 July 2023